

## Replicated user comment<sup>1</sup>

*Replicated join pattern (Challenge 6<sub>b</sub>)*

For this exercise you have to prepare a report including:

- Your data collections comment describing what they represent.
- Your source codes: a file precising in a comment the compiling and execution commands.
- Propose a test battery and report execution measures.
- Compress everything and sent a .zip or a .tar to [genoveva.vargas@gmail.com](mailto:genoveva.vargas@gmail.com)

### 1.1 Problem statement without filter

Problem: Given a small set of user information and a large set of comments, enrich the comments with user information data.

### 1.2 Implementation

Look at page 122 of the book “Map Reduce design patterns” and see the proposed Map code. Prepare a data collection of your choice and implement the solution. As in previous challenges, you can use StackOverflow or any other source for generating your collection. Program the first part of the example.

- Describe the general principle of the solution proposed by the book.
- Explain the purpose and necessity of the HashMap in the map method. Which are the implications of using it?
- Implement and compare intuitively the other joins.
- Explain the raisons for the absence of a reducer.
- Prepare collections of different sizes to run your tests trying to get to the limits of your solution.
- Make comparisons. Do not hesitate to prepare graphics.

---

<sup>1</sup> This challenge is an example proposed in the book MapReduce design patterns, pp. 121.